

# Linguistic patterns in Google search behaviour

Vlad Mackevic, Annina Heini, Krzysztof Kredens (Centre for Forensic Text Analysis)

### Context

Members of the Centre for Forensic Text Analysis were approached to assist in two murder cases involving internet search engine data. Police investigators were keen to find out which of a closed set of suspects with access to a seized device had used Google to research topics associated with the murders. Records of the searches conducted by the suspects prior to the murders were extracted from their mobile phones and provided for analysis. However, with no existing research on individual search engine behaviour, the authorship attribution task was difficult and the findings only had limited value for the investigation.

### Data collection

112 individuals completed a scenario-based task. As PAs to a fictitious CEO, they were asked to research a number of items for the CEO's business trip:

Dear PA,

The CEO is travelling to Helsinki for a 2-day conference from 2.-4. September 2021. Please could you find the following itinerary items for her trip:

1. **A central hotel, budget around £200 per night. Note that she likes to stay close to the water**
2. A café to have breakfast at
3. A museum to learn about the city's history
4. A restaurant that serves a typical national dish. Note that she is allergic to seafood
5. A shop to buy a typical Helsinki souvenir for her children
6. Something to do in the evening. Note that she loves live music but does not like opera

Participants' Google search behaviour was captured using screen recording and their written input transcribed manually.

### Findings

Google search behaviour tends to differ across our participants and could be idiolectal. As an illustration, consider the range of variation observed in the search strings for item 1, **a central hotel, budget around £200 per night. Note that she likes to stay close to water.**

Part.	Search string
57	central Helsinki Hotels near water
53	central hotel in Helsinki near sea
53	central hotel in Helsinki sea front
17	Hellsinki (sic) hotel central close to water
30	Helsinki beach hotels
47	Helsinki central hotels by water
92	Helsinki central hotels water view
49	Helsinki hotel bookings central close to waterfront
8	Helsinki hotel near water
95	Helsinki hotel sea view
45	Helsinki hotels by the water
88	Helsinki hotels close to water
20	Helsinki hotels near riverside
82	Helsinki hotels near the beach
95	Helsinki hotels near the sea
21	Helsinki, central hotel close to water
52	hietarana (sic) beach hotel nearby Helsinki
23	hotel in Helsinki close to water
74	hotels in central Helsinki near river
74	hotels in central Helsinki near river £200
51	hotels in Helsinki by water
50	hotels in Helsinki Finland city centre by the water
99	hotels in Helsinki Finland city centre near water
84	hotels in Helsinki near beaches
74	hotels in Helsinki with waterfront
97	hotels in hHelsinki (sic) near the water
74	hotels near river Helsinki
74	hotels near river Helsinki booking.com
26	hotels near water in Helsinki

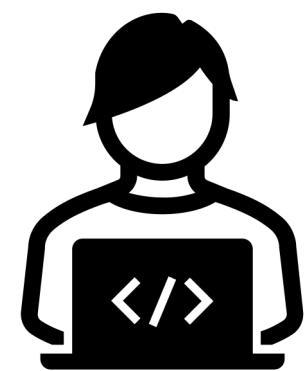
Table 1: Participant numbers and search strings for item 1 of task

### Variation

The data in Table 1 show **high levels of variation across participants**. They use both singular and plural forms (*hotel* vs. *hotels*), include and omit articles (*by water* vs. *by the water*), and list specific bodies of water (*river, beach, etc.*). Word order is also varied, e.g. the word *Helsinki* occurs at the start of the search string or in the middle/at the end with a preposition (*in Helsinki*).

### Within-author consistency

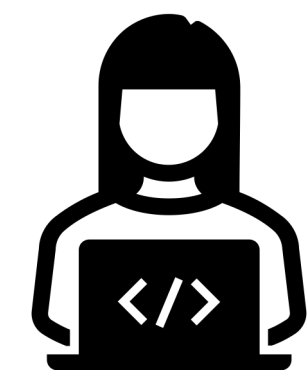
The data overall show **low levels of consistency within participants**. The majority (76%) are not consistent syntagmatically in their Google search behaviour. Consider the placement of 'Helsinki' in these search strings by the same user:



Participant 113

Clarion Hotel **Helsinki**  
**helsinki** souvenir shop  
live music **helsinki**  
cafes **helsinki** central

Only a few participants (14%) show clear signs of consistency, e.g. by always placing the word 'Helsinki' at the start of their search string:



Participant 39

**helsinki** live music bar  
**helsinki** central breakfast cafe/restaurant  
**helsinki** finnish food restaurant  
**helsinki** things to do

Our findings suggest internet users are likely to have individual patterns of Google search behaviour. These patterns seem to be more complex than expected and the sites of variation include singular vs. plural forms, word order, word choice, use of prepositions, use of questions, and query specificity.

